AUDITION AND VISION IN VIRTUAL REALITY

Thomas Piantanida, Principal Scientist
Medical Technology Laboratory

SRI Project 5729

Prepared for:

19950922 062

OFFICE OF THE UNDER SECRETARY OF DEFENSE (ACQUISITION)
DEFENSE TECHNICAL INFORMATION CENTER
CAMERON STATION
ALEXANDRIA, VIRGINIA 22304-6145

Oct 25, 1994

IN REPLY
REFER TO

DTIC-OCC

SUBJECT: Distribution Statements on Technical Documents

TO:
Office of the Chief of Naval Research
800 north Quincy Street
Arlington, VA 22217-5000
Code

1. Reference: DoD Directive 5230.24, Distribution Statements on Technical Documents, 18 Mar 87.

2. The Defense Technical Information Center received the enclosed report (referenced below) which is not marked in accordance with the above reference.

Contract No. N00014-94-C-0097
Audition and vision in Virtual Reality

3. We request the appropriate distribution statement be assigned and the report returned to DTIC within 5 working days.

4. Approved distribution statements are listed on the reverse of this letter. If you have any questions regarding these statements, call DTIC's Cataloging Branch, (703) 274-6837.

FOR THE ADMINISTRATOR:

1 Encl

for GOPALAKRISHNAN NAIR
Chief, Cataloging Branch

FL-171
Jul 93

AUDITION AND VISION
IN VIRTUAL REALITY

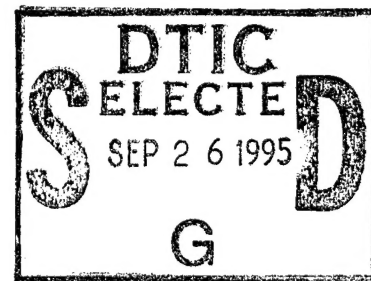Thomas Piantanida, Principal Scientist
Medical Technology Laboratory

SRI Project 5729

Prepared for:

Cognitive and Neural S&T Division
ONR 342
Office of Naval Research
800 North Quincy Street
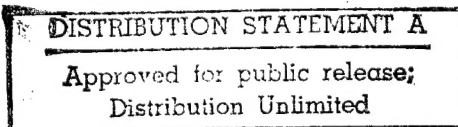Arlington, VA   22217-5660

Attn: Harold L. Hawkins

Contract N00014-94-C-0097

Accesion For		
NTIS CRA&I		☒
DTIC TAB		☐
U..announced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

Approved:

Ajit Shah, Director
Medical Technology Laboratory

Donald L. Nielson, Vice President
Computing and Engineering Sciences Division

SRI International   333 Ravenswood Avenue • Menlo Park, CA 94025-3493 • (415) 326-6200 • FAX: (415) 326-5512 • Telex: 334486

CONTENTS

1. INTRODUCTION

SRI International has completed the first six-month period of a program designed to provide critical information about the most effective use of visual and acoustic icons to improve intelligibility and comprehension of data representations in virtual environments. Major accomplishments during this period include configuring the virtual-reality (VR) system for conducting psychophysical experiments and performing initial target-detection studies, using visual and acoustic spatial-orientation cues. These pilot studies revealed technological and psychophysical issues that required us to modify and develop experimental protocols for conducting psychophysical studies in a virtual environment (VE).

Preliminary observation indicates that traditional metrics, such as time-to-target detection, may be inappropriate performance measures for psychophysical experiments conducted in virtual environments because of current technological limitations of VR systems. We have adopted a path-length-to-detection metric that is independent of head/body-movement velocity.

We also found that cues to absolute orientation—chromatic grids and localized acoustic icons—may diminish target-detection performance whether measured in terms of time to detection or path length to detection. A possible explanation is that absolute orientation cues, which tell the subject the direction he is facing at any given moment, may hamper task performance by increasing cognitive load.

2. VIRTUAL-REALITY SYSTEM CONFIGURATION

We assembled a combination of hardware and software components to produce a virtual world that we could use to test the effects of various visual and acoustic parameters on human performance in the VE. The components of the system used to generate the VE, as well as the contents of the VE, are described below.

2.1 HARDWARE

In our experiments, a subject's main interfaces to the field-of-view (FOV) system were a VPL EyePhone* providing a wide field-of-view, stereoscopic image; headphones playing binaural sound; head and hand tracking via a Polhemus 3Space 6-degree-of-freedom (DOF) tracker; and a Virtex Cyber Glove measuring the joint angles of the subject's fingers.

---

* All product names mentioned in this report are the trademarks of their respective holders.

The input devices (glove and tracker) fed into the serial ports of a Macintosh IIfx running the simulation with VPL's Body Electric software. To update the visuals, the Macintosh sent position information to two Silicon Graphics Crimson VGXs, running VPL's Isaac software, which rendered the scene at 20 Hz. The video from these workstations was fed through RGB to NTSC converters and then into the liquid crystal displays (LCDs) in the EyePhone for display.

To produce localized or binaural sound, the Macintosh sent MIDI commands to a Peavy SP synthesizer, which produced the sounds on four different channels. These four channels were then fed through a Crystal River Engineering Convolvotron that spatialized the monaural sounds using location information received over a serial line from the Macintosh. The localized sound signal was sent over a stereo audio cable to the headphones in the EyePhone.

Position information from the 6-DOF trackers was used to change the viewpoint of the scene rendered by the Silicon Graphics workstations. The Cyber Glove measured joint angles of the fingers to detect gestures such as fist or open hand.

2.2   SOFTWARE

The simple world geometry used in the experiments was modeled using Swivel 3D. VPL's Body Electric, a visual programming language, was used to create the experiment control software. Some of the software used in this study had already been created for earlier versions of FOV experiments using Body Electric, so it was more efficient to expand and refine this software, rather than starting anew. At the beginning of each experimental trial, the software moved the target to a random azimuth within ±45 degrees elevation, but not in the user's current FOV. Each distractor was also moved to a new position. This movement was random, except that the distractor could not overlap another distractor or the target.

The duration of each trial, from new target being displayed, to target-detection signal, was recorded in a spreadsheet for later analysis. The path length in degrees of the user's line of sight was also recorded for evaluation of different search strategies used with different FOVs.

2.3   VIRTUAL WORLDS

The virtual-world geometry consisted of a uniform, white, subject-centered sphere with target, distractor, and orientation elements affixed to the inner surface. Specific elements were presented on the inner surface, with retinal disparities consistent with a 60-foot-diameter sphere. The virtual sphere was fixed in position in real-world coordinates. Subjects had a compelling impression of being centered within a large, spherical chamber. As many as four sound sources with different or identical, simple or complex waveforms could be localized at prescribed latitudes and longitudes on the sphere.

The unrestricted FOV of the virtual sphere (imposed by the geometry of the EyePhone) was approximately 100 degrees horizontal and 60 degrees vertical. The subject viewed the inner surface of the sphere through one of several virtual apertures subtending 14, 28, 41, or 53 degrees of visual angle. The mask surrounding the aperture was identical in color to the inner surface of the sphere. Therefore, when no elements on the inner surface of the sphere were visible through the aperture, the aperture mask could not be distinguished from the sphere's inner surface—that is, the perimeter of the aperture was not detectable. Previous experiments in our

laboratory suggested that a contrasting aperture mask, such as a black mask and white inner sphere surface, increased the probability of the subject's experiencing simulator sickness, so zero-contrast aperture masks were used throughout the study.

In the experiments conducted thus far, the subject's task was to detect the presence of a target. In each of the virtual worlds, the target consisted of a red square positioned on the inner surface of the sphere at any azimuth and at any elevation within ±45 degrees. The red square subtended approximately 6 degrees of visual angle on a side. Similarly sized distractors, usually four 6-degree-diameter red circles and four 6-degree-square green squares were randomly distributed over the 32,400-square-degree (360-degree azimuth, 90-degree elevation) inner surface, with the proviso that a distractor could never occlude the target. There were never any blank (i.e., no-target) trials.

The subject signaled that the target had been detected by clenching his fist. This gesture was recognized by the computer as a signal to record the four-space coordinates of the target (yaw, pitch, radius [constant], and time) and to initiate a new trial in which target and distractors were presented in new locations. On most trials, it was necessary for the subject to move his head, and often his body, to locate the target object. Pitch and yaw of the subject's head were measured continuously, and these measurements had a twofold purpose: to eliminate false alarms (such as when the subject indicated that he had detected the target, but the target was not present in his FOV), and to calculate the path lengths between successive target detections. Time between successive target detections was also measured.

An achromatic spherical grid pattern could be superimposed on the inner surface of the sphere to provide the subject with orientation cues. Our previous studies indicated that in the absence of orienting cues, cognitive errors in remapping the virtual world after each body movement accumulated. Eventually, the subject believed that he was facing north, for example, when he was actually facing east, west, or south.

The achromatic spherical grid consisted of major black lines of longitude every 90 degrees, that subtended about 6 degrees of visual angle; minor black lines of longitude every 30 degrees, that subtended about 3 degrees of visual angle; a major black latitude line at the equator, subtending about 6 degrees of visual angle; and minor black latitude lines every 15 degrees, subtending about 3 degrees of visual angle. The grid was affixed to the inner surface of the sphere in real-world coordinates.

Similarly, the experimenter could substitute a chromatic grid pattern to provide additional orientation cues. The chromatic grid consisted of the achromatic grid with discriminable chromatic bars substituted for the major longitudinal lines. Thus, as the subject scanned the inner surface of the sphere, he would see at 90-degree intervals, a red, blue, green, or brown longitude line at the major compass points. All minor longitude lines and all latitude lines were identical to the achromatic grid.

The chromatic grid provided the subject with an absolute frame of reference, as the color of the grid informed the subject that he was facing a particular compass point. The achromatic grid, however, provided only a relative frame of reference—that is, the subject could tell how far he had rotated, but not unambiguously the direction he was facing.

The subject was provided with localized acoustic icons that could be used for orientation cues. As many as four icons could be presented, each with a distinguishable acoustic signature. When four acoustic icons were used, the acoustic array consisted of a subject-centered regular tetrahedron with one acoustic source at the north pole of the sphere. Single acoustic icons could be positioned anywhere in the sphere, but practically, it was advantageous to position the acoustic source on the equator of the sphere.

3. PROTOCOLS

A typical experimental session consisted of an initial period during which the subject donned the EyePhone and Cyber Glove, and the experimenter calibrated the EyePhone for the subject's interpupillary distance and the Cyber Glove for the subject's hand. The experimenter then selected the appropriate virtual-world parameters (e.g., aperture size, chromatic grid/ achromatic grid/no grid, acoustic icons), initialized the experiment, and presented the first trial. Thirty-three trials were run as a block. Data were collected for the last thirty-one trials. At the conclusion of a block, a tone sounded, indicating to the subject that he had completed the block. The subject then took a brief break while the experimenter changed stimulus parameters and initiated another block of trials.

3.1 TIME TO DETECTION

The performance metric used in our previous experiments was time-to-target detection (TD). Detection time was recorded as the interval between successive target detections, as indicated by the subject's hand gestures. This measure was effective in the earlier experiments, but became problematical in the present study. Subjective reports indicated that the presence of orientation grids caused subjects to reduce their speed of head rotation to avoid simulator sickness. A reduction of head-movement speed artificially increased detection time for some subjects on trials that included visual orientation cues, thereby confounding the effects of those cues. Fortunately, data that allowed us to calculate path length to detection (PD) were available from the tracker, although we continued to measure TD on each trial.

3.2 PATH LENGTH TO DETECTION

Head-tracker data consisted of yaw, pitch, roll, and X, Y, and Z position signals. Pitch and yaw signals allowed us to calculate the orientation of the subject's head thirty times per second, and from these data, we could map the path of the subject's eyes in the virtual world. Using these data, we were able to calculate the path length between successive target detections in cumulative degrees of pitch and yaw. This metric, path length to detection (PD), became our primary performance measure because it was not contaminated by changes in head-movement velocity. Usually, with a small FOV, scanning search strategies that incorporate long path lengths are adopted, whereas with a large FOV, a random, short-path-length search is sufficient to find the target.

4. PRELIMINARY RESULTS

We have conducted a limited number of pilot studies to identify appropriate stimulus parameters for subsequent study and to evaluate the effectiveness of our experimental protocols in overcoming the technological limitations of our VE system. Reported below are preliminary results obtained with a limited number of subjects in experiments designed as much to evaluate our developing protocols as to reveal the effects of stimulus parameters.

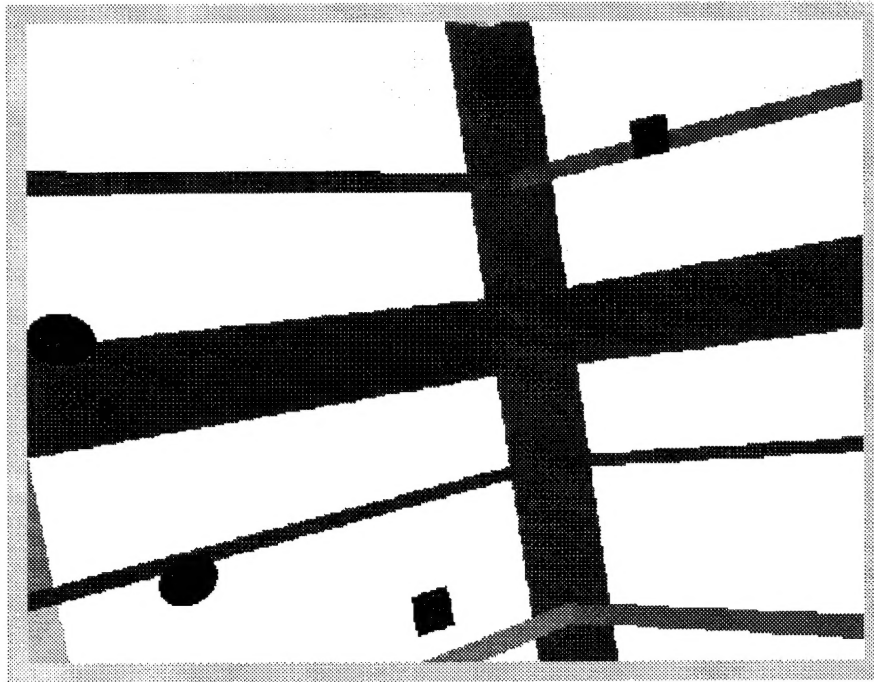## 4.1 VISUAL ORIENTATION CUES

As noted above, we found conflicting results of TD and PD under some stimulus conditions. TD provided ambiguous across-subjects comparisons under conditions in which a visual orientation cue (grid) was present, but PD did not. TD measures suggested that orientation grids could either improve performance (i.e., reduce detection time) or degrade performance. PD measures, however, indicated that the orientation grids consistently improve target detection, probably by reducing redundant searching. That is, in the absence of orientation cues, subjects may unwittingly research previously searched areas of the inner surface of the sphere, but the cognitive map of the inner surface of the sphere provided by the orientation cues allowed them to avoid searching previously searched areas.

There is some preliminary indication that chromatic grids were somewhat worse than achromatic grids at providing orientation cues, although data are sparse on this question. One possible explanation is that chromatic grids impose a greater cognitive load on the subjects by requiring them to constantly correlate color and compass points, which, in fact does not aid in the detection task.

The accompanying figure compares a target view with and without an aperture.

4.2 DISTRACTORS

We have not yet systematically assessed the effect of distractors. As we expected, however, color is not subjectively as distracting as shape. Because distractors were widely separated in color space, they were capable of being processed preattentively—that is, green distractors were rejected without further scrutiny. Same-color distractors—such as red circles—required scrutiny to determine that they were not, in fact, the target. Resolution limitations of the EyePhone may have contributed to the ambiguity of distractor shape. In subsequent studies, we will need to explore other distractor shapes and orientations that may be less affected by EyePhone resolution.

(a) View of target and three distractors with grid and no aperture
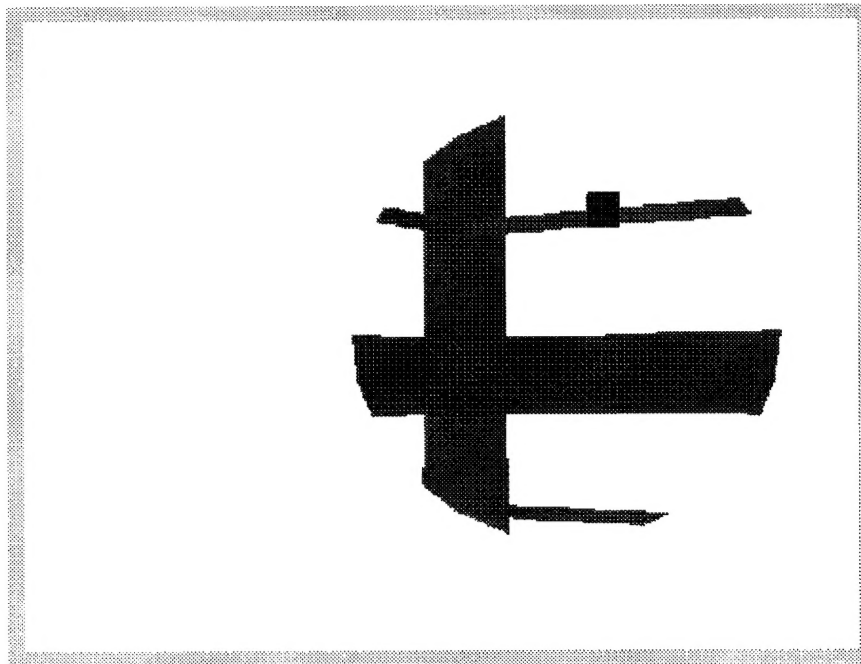
(b) View of target with grid and 28° aperture

Comparison of target view with and without aperture

4.3   ACOUSTIC ORIENTATION CUES

Acoustic cues were, in general, unimpressive. Initial observations suggest that they increase both TD and PD, even in the presence of visual orientation cues. The major factor seems to be auditory localization, which was generally poor. Poor auditory localization was due, perhaps, to technological limitations, to lack of visual capture in the virtual world we used, or to subject anatomy and auditory psychophysics. Preliminary results suggest that, contrary to our original expectation, a single acoustic icon may be more effective at providing spatial orientation cues, at least where only yaw and pitch (not acoustic pitch) are the parameters of interest.

Our initial observations also suggest that acoustic signature is an important parameter in auditory localization. Broad-spectrum acoustic sources with abrupt transients appear to provide better localization cues than pure tones. Our typical acoustic orientation cue consisted of a localized stream of clicks. Further work is necessary to optimize acoustic icons as orientation cues.

One of the more significant discoveries of this study is that in virtual environments, visual reference cues, such as chromatic and achromatic grids, may not improve target-detection performance as measured by a time-to-detection metric. Thus, because of current technological limitations of VR systems, traditional metrics, such as time-to-target detection, may not be appropriate measures of task performance in virtual environments.

5.   RESULTING DATA

The data obtained in our preliminary studies show large scatter and typically a few outliers, resulting in unusually large standard deviations. Thus, the data appear not to be distributed normally, which prompts us to adopt a nonparametric statistic, such as the Sign Test or, more likely, the Wilcoxon Matched-Pairs Signed-Ranks Test. Initial analyses using the less powerful Sign Test support the tentative conclusion that the path lengths of scanning strategies are shorter in virtual worlds that provide visual orientation cues.

6.   FUTURE YEAR-1 WORK

With much of the protocol-development and system-development aspects of the FOV experiments behind us, we are prepared to validate our initial observations with data acquired from a larger number of subjects. We expect to explore the following areas:

- Relative efficacy of chromatic and achromatic spatial-orientation grids as a function of aperture size

- Effects of distractor parameters (e.g., shape, orientation) on target detection

- Effects of visual capture of acoustic icons on spatial orientation

- Effects of acoustic signatures on spatial orientation

To enhance interpretability of our experimental results, it will be necessary to optimize the pertinent visual and acoustic parameters of the virtual world. Understanding the effects of such stimulus parameters as color, orientation, timbre, and localization will enable us to generalize our results to other virtual- and real-world situations.

In addition to optimizing virtual-world stimulus parameters, the major thrust of experiments proposed for the remainder of the first year will be to elucidate the effects of field size on detection-task performance. Our results will suggest the requisite field size and spatial orientation cues for performing search tasks in virtual worlds that simulate such diverse environments as the ocean floor, or a crippled nuclear reactor, or the lumen of the stomach.

7. DISSEMINATION OF INFORMATION

Papers describing the results of this recently initiated project have not yet been submitted or accepted for publication in refereed journals, and no technical reports or patents have been issued.

The Principal Investigator (PI) has been asked to develop an outline for a chapter that will describe work on the perceptual issues of VR being conducted in our laboratories and in other laboratories. This chapter will be part of a book, commissioned by McGraw-Hill and edited by Kirk Moffitt and Jim Melzer of Kaiser Electronics, on the requirements of head-mounted displays. The work being conducted under ONR support will be an essential element of this book. The book is expected to become an important reference source for manufacturers of head-mounted displays, as well as for end users of such displays in industry, the military, and education. Initial research findings to be reported in the book have already influenced the PI's approach to developing VR-based intelligent tutoring systems.

The PI was invited to speak at The Virtual Reality Group in San Francisco in July, where he presented a lecture on visual factors in VR. He has been invited to speak on perceptual issues in VR at Ausi VR '94 in Melbourne, Australia, in October 1994, and at the Interface to Real and Virtual Worlds Conference in Montpelier, France, in May 1995. The PI has also been invited to sit on a panel that will discuss visual factors in VR at Virtual Reality Expo '94 in New York in November 1994.

Our laboratory staff includes one predoctoral student, International Fellow Denis Amselem from Grenoble University, who has been developing a hand-held interface to virtual worlds that uses the results of the ONR-sponsored program to enhance user comfort and cognition. He is also part of the laboratory team that is developing algorithms of collaborative

VR systems shared by multiple users. The information we have acquired through the ONR-sponsored project has provided insights into methods of reducing bandwidth requirements of collaborative VR systems at little or no cost to user cognition.

We are currently in discussion with Stanford University to develop a program that will enable graduate students in a combined Computer Science and Cognitive Psychology Program to conduct research in our laboratory. An initial candidate for that program has been selected, and he may start to work in our laboratory within the next few months.

No equipment costing more than $1,000 has been purchased on this ONR grant.